LINKAGE DISEQUILIBRIUM BETWEEN TWO SEGREGATING NUCLEOTIDE SITES UNDER THE STEADY FLUX OF MUTATIONS IN A FINITE POPULATION[1]

TOMOKO OHTA AND MOTOO KIMURA

National Institute of Genetics, Mishima, Japan

Received July 31, 1970

IN our previous reports (OHTA and KIMURA 1969a, b, 1970), we have studied linkage disequilibrium, that is, nonrandom association of genes between loci, caused by random frequency drift in finite populations. We have considered the situation in which a stationary distribution is reached under recurrent mutation or overdominance. We have also considered the case in which genetic variability decays each generation due to random sampling of gametes.

The present paper is an extension of the work of KIMURA (1969a) who studied the number of heterozygous nucleotide sites under the steady flux of molecular mutations in a finite population. Here, we intend to study the amount of linkage disequilibrium between two segregating sites using the same model; it assumes that the total number of nucleotide sites making up the genome is so large and the mutation rate per site is so low that whenever a mutant appears, it represents a mutation at a previously homoallelic site, that is, a site in which no mutants are currently segregating in the population.

BASIC THEORY

Consider a random mating diploid population of actual size N and effective size N_e. We assume that each generation mutations occur in v_m sites distributed throughout the population.

Since each mutant becomes fixed in the population or lost from it within a finite length of time, if mutations continue to occur at a constant rate over many generations, a steady state will be reached with respect to the frequency distribution of mutants among different sites, provided that we restrict our consideration to only those sites in which mutant forms are segregating.

In the following treatment, we shall consider two segregating nucleotide sites with the recombination fraction c between them. Let X_1 be the frequency of chromosomes having no mutants at both sites, X_2 and X_3 be the frequencies of chromosomes having a mutant at the first and the second sites, respectively, and X_4 be the frequency of chromosomes having mutants at both sites. Then, $x=X_2+X_4$ is the frequency of the mutant at the first site, $y=X_3+X_4$ is that at the second site, and $D=X_1X_4-X_2X_3$ is the index of linkage disequilibrium.

[1] Contribution No. 782 from the National Institute of Genetics, Mishima, Shizuoka-ken 411 Japan. Aided in part by a Grant-in-Aid from the Ministry of Education, Japan.

Let $\Phi(x,y,D)$ be the steady flux distribution involving two nucleotide sites such that $\Phi(x,y,D)\mathrm{d}x\mathrm{d}y\mathrm{d}D$ is the expected number of the pairs of sites having mutant frequencies and the disequilibrium index within the intervals $(x,x+\mathrm{d}x)$, $(y,y+\mathrm{d}y)$, and $(D,D+\mathrm{d}D)$. This includes all pairs whose distance apart corresponds to a recombination fraction c. We assume that mutants are simultaneously segregating at both sites $(0<x<1, 0<y<1)$.

As shown in the APPENDIX, if $f(x,y,D)$ is a function (polynomial) of x,y, and D, we have, with a steady flux of mutations,

$$E\{L(f)\} + \Delta_{mut}E(f) = 0. \tag{1}$$

In this equation L is the differential operator for the diffusion process involving the two sites, and if we assume that the molecular mutants are selectively neutral,

$$L = \frac{x(1-x)}{4N_e}\frac{\partial^2}{\partial x^2} + \frac{y(1-y)}{4N_e}\frac{\partial^2}{\partial y^2} + \frac{D}{2N_e}\frac{\partial^2}{\partial x\partial y} + \frac{D(1-2x)}{2N_e}\frac{\partial^2}{\partial x\partial D}$$
$$+ \frac{D(1-2y)}{2N_e}\frac{\partial^2}{\partial y\partial D} + \frac{1}{4N_e}[xy(1-x)(1-y)+D(1-2x)(1-2y)-D^2]\frac{\partial^2}{\partial D^2}$$
$$- \left(\frac{1}{2N_e}+c\right)D\frac{\partial}{\partial D}\ . \tag{2}$$

This operator is equivalent to L_B in formula (12) of OHTA and KIMURA (1969b) except that L here does not contain terms involving the effect of mutation. In the present model, considering nucleotide sites rather than conventional genetic loci, mutation is essentially irreversible and the effect of mutation is represented by the term $\Delta_{mut}E(f)$ in equation (1). Specifically, $\Delta_{mut}E(f)$ represents the contribution made each generation by new mutations to $E(f)$, where E is the operator for taking the expectation with respect to the steady flux distribution, that is,

$$E(f) = \iiint f(x,y,D)\Phi(x,y,D)\mathrm{d}x\mathrm{d}y\mathrm{d}D,$$

in which the integral is over $0<x<1$, $0<y<1$, $-\frac{1}{4}\leqq D\leqq+\frac{1}{4}$.

In choosing f we must keep in mind that equation (1) is valid only for f such that $f(x,y,D)\Phi(x,y,D)$ vanishes at $x=0$, $x=1$, $y=0$, and $y=1$; in other words, $f\Phi$ must be zero on the periphery of the square $0\leqq x\leqq 1$, $0\leqq y\leqq 1$. Note that on the periphery D is also zero.

First, let $f=xy(1-x)(1-y)$ in equation (1), then we have

$$E\{L(f)\} = \frac{1}{2N_e}(-2X+Y)\ , \tag{3}$$

where $X=E\{xy(1-x)(1-y)\}$ and $Y=E\{D(1-2x)(1-2y)\}$. To determine $\Delta_{mut}E(f)$ in equation (1), let ν_s be the number of pairs of nucleotide sites that start segregating simultaneously in the entire population each generation, considering only those pairs of sites that are separated by a distance corresponding to a recombination fraction c. We assume that simultaneous segregation always starts from the situation in which one of the sites is already segregating while a new mutant is just added to the other site. Thus for $f=x(1-x)y(1-y)$, we have

$$\Delta_{mut}E(f) = \nu_s\overline{x(1-x)}p(1-p),$$

where $\overline{x(1-x)}$ is the average value of $x(1-x)$ among segregating sites, and p is the frequency of mutants at the time of occurrence.

If we denote by $I_1(p)$ the total number of segregating sites in the population and by $E\{x(1-x)\}$ the sum of $x(1-x)$ over all these sites, then

$$\overline{x(1-x)} = E\{x(1-x)\}/I_1(p),$$

and as shown by KIMURA (1969a),

$$E\{x(1-x)\} = 2N_e v_m p(1-p) \tag{4}$$

and

$$I_1(p) = -4N_e v_m\{p \log_e p + (1-p) \log_e(1-p)\}.$$

An independent derivation of (4) is given in the APPENDIX (see formula A7).

We note here that we may put $p=1/(2N)$ in the above expressions because, in our model, the mutation rate per site is so low that each mutant is likely to be represented only once at the moment of appearance. Thus, if we write

$$\Delta_{mut}E(f) = K$$

for $f=x(1-x)y(1-y)$, we have

$$K = v_s\overline{x(1-x)}/(2N) = v_s/[4N(\log_e 2N+1)] \tag{5}$$

approximately.

Next, letting $f=D(1-2x)(1-2y)$ in (1), we obtain

$$E\{L(f)\} = -\frac{1}{2N_e}\{(5+2N_e c)Y - 4Z\},$$

where $Z = E\{D^2\}$. In determining $\Delta_{mut}E(f)$ for this case, we note that when a mutant is just introduced into the second site, the mutant appears either on a chromosome carrying a mutant in the first site in which case $X_4 = 1/2N$, $X_3 = 0$, $X_2 = x - 1/(2N)$, $X_1 = 1 - x$, or on a chromosome carrying a nonmutant in the first site in which case $X_4 = 0$, $X_3 = 1/2N$, $X_2 = x$, $X_1 = 1 - x - 1/2N$, where x is the frequency of the mutant in the first site.

The frequencies of these alternative events are x and $1 - x$, respectively. Therefore,

$$\Delta_{mut}E(f) \propto E\Big[x\Big(\frac{1-x}{2N}\Big)(1-2x)\Big(1-\frac{2}{2N}\Big) + (1-x)\Big(-\frac{x}{2N}\Big)(1-2x)\Big(1-\frac{2}{2N}\Big)\Big] = 0.$$

Finally, let $f = D^2$ in (1), then we get

$$E\{L(f)\} = \frac{1}{2N_e}\{X + Y - (3+4N_e c)Z\},$$

and

$$\Delta_{mut}E(f) = v_s\overline{x(1-x)}/(2N)^2 \approx K/(2N),$$

because for a particular value of x, the contribution of a new mutant to $E(D^2)$ is

$$v_s\Big[x\Big(\frac{1-x}{2N}\Big)^2 + (1-x)\Big(\frac{-x}{2N}\Big)^2\Big] = v_s x(1-x)/(2N)^2.$$

Therefore, we have a set of equations for X, Y, and Z.

$$\begin{aligned} -2X+Y &= -2N_eK \\ -(5+2N_ec)Y+4Z &= 0 \\ X+Y-(3+4N_ec)Z &= -2N_eK/(2N) \end{aligned} \tag{6}$$

Solving this we find

$$\begin{aligned} X &= E\{x(1-x)y(1-y)\} = N_eK(11+26R+8R^2+2/N)/(9+26R+8R^2) \\ Y &= E\{D(1-2x)(1-2y)\} = 4N_eK(1+1/N)/(9+26R+8R^2) \\ Z &= E\{D^2\} = N_eK(1+1/N)(5+2R)/(9+26R+8R^2), \end{aligned} \tag{7}$$

where $R = N_ec$.

To express the degree of linkage disequilibrium between the two nucleotide sites, we use the quantity,

$$\sigma_d^2 = \frac{Z}{X} = \frac{E\{D^2\}}{E\{x(1-x)y(1-y)\}}, \tag{8}$$

where σ_d is the *standard linkage deviation* (OHTA and KIMURA 1969b). The quantity $D^2/[x(1-x)y(1-y)]$ has been used by statisticians as a measure of degree of association and is called the *mean square contingency coefficient* (see KENDALL 1948). Our σ_d^2 is closely related, being the ratio of the expected value of the numerator to that of the denominator. Unless the mutant frequencies at one or both sites take the extreme values near 0 or 1, σ_d^2 is approximately equal to the expected value of the mean square contingency coefficient.

From (7) we obtain

$$\sigma_d^2 = \frac{(5+2R)(1+1/N)}{11+26R+8R^2+2/N} \approx \frac{5+2R}{11+26R+8R^2} \tag{9}$$

where $R = N_ec$.

Thus if N_ec is large, σ_d^2 is approximately equal to $1/(4N_ec)$, while if N_ec is much smaller than unity, σ_d^2 is approximately 5/11.

Note that $E(D) = 0$, namely, D has the mean zero, as may be seen by setting $f = D$ and $\Delta_{mut}E(f) = 0$ in (1). However, for any finite population, D is likely to deviate from this theoretical mean, and σ_d is a more appropriate measure of the amount of linkage disequilibrium.

MONTE CARLO EXPERIMENTS

In order to check the validity of these theoretical predictions, several Monte Carlo experiments were performed. It is desirable to simulate as closely as possible a natural population of organisms having a very large number of nucleotide sites in its genome with a steady flux of mutations occurring over many generations. However, we used a simpler model having only two sites, corresponding to the two sites treated by the method of this paper. Each site is supplied with a new mutant as soon as it becomes homoallelic. In this model we cannot control mutation rate, although for the present purpose, the simulation may be used for checking formula (9) giving the squared standard linkage deviation. Also, it may be used to compare σ_d with the contingency coefficient between the two nucleotide sites.

TABLE 1

$N_e c$	σ_d^2			
	Theoretical value (formula 9)	Monte Carlo results		
		N_e=100	N_e=200	Mean
0.0	0.4545	0.7380	0.2933	0.5157
0.1	0.3824	0.7056	0.6888	0.6972
0.2	0.3273	0.4571	0.4637	0.3922
0.3	0.2872	0.3903	0.2930	0.3414
0.4	0.2557	0.4526	0.2989	0.3758
0.5	0.2308	0.1890	0.2532	0.2211
0.6	0.2102	0.2008	0.1197	0.1603
0.7	0.1932	0.1823	0.1348	0.1586
0.8	0.1788	0.1700	0.2606	0.2153
0.9	0.1663	0.1273	0.0053	0.0663
1.0	0.1555	0.1334	0.2554	0.1944
2.0	0.0947	0.1208	0.0858	0.1033
3.0	0.0683	0.0622	0.0768	0.0695
4.0	0.0535	0.0391	0.0380	0.0386
5.0	0.0440	0.0444	0.0506	0.0475
6.0	0.0374	0.0319	0.0310	0.0315
7.0	0.0325	0.0313	0.0273	0.0293
8.0	0.0287	0.0263	0.0231	0.0247
9.0	0.0258	0.0238	0.0211	0.0225
10.0	0.0233	0.0197	0.0192	0.0195

Results of Monte Carlo experiments performed to check equation (9) on the squared standard linkage deviation between two segregating sites under the steady flux of molecular mutations in a finite population. Each experimental value is the average over 10,000 generations. N_e stands for the effective population number and c the recombination fraction between the two segregating nucleotide sites.

The procedure of the experiment was as follows. In the first generation, each site contains one mutant; crossing over and zygote formation are performed deterministically; sampling of N zygotes for the parents of the next generation are carried out as follows using pseudorandom numbers with uniform distribution in the interval [0,1] (RAND 20 in TOSBAC 3400). Let f_i be the frequency of the ith genotype ($i = 1,2,\ldots,10$). Then we pick out an individual of the ith genotype if a random number happens to lie between $\sum_{j=0}^{i-1} f_j$ and $\sum_{j=0}^{i} f_j$, where we set $f_0 = 0$. The procedure used here is essentially the same as the one used by OHTA (1968) except for mutation production. A new mutation is supplied whenever a locus becomes homoallelic, regardless of whether the previous mutant was lost or fixed. In Table 1, values of σ_d^2 obtained from the experiment are compared with the corresponding theoretical values derived from equation (9). The experiments were performed assuming two levels of the effective population number, $N_e = 100$ and 200. The recombination fraction ranges from 0 to $N_e c = 10$. The values listed are the averages over 10,000 generations.

As seen from the table, agreement between experimental and theoretical results appears to be satisfactory.

DISCUSSION

In discussing the biological implication of the above results, we must keep in mind that we are here concerned with nonrandom association of mutants between two nucleotide sites, both of which are *simultaneously segregating* in the population. This differs from the nonrandom association reported by JOSSE, KAISER and KORNBERG (1961) for base composition of adjacent nucleotide sites (nearest neighbor). Their analysis involves the overall base composition, most of which is from nonsegregating sites. Furthermore, our analysis considers nonrandom associations between nucleotide pairs, considered as individual pairs, but with an average value D not differing from zero, whereas they found significant deviations from $D = 0$.

The most likely explanation of the nearest-neighbor association is that the mutation rate at a nucleotide is somehow influenced by the nucleotide at the adjacent site. The alternative, that the nonrandomness is due to selection favoring different amino acids is more difficult to understand in view of the fact that the same paired nucleotide sequences occur in the codons of many amino acids and many different amino acids are used in each polypeptide. If a majority of nucleotide substitutions in evolution are the result of chance fixation of molecular mutants through random frequency drift as suggested by KIMURA (1968, 1969b), KING and JUKES (1969), and CROW (1969), nonrandomness of base arrangements between adjacent sites is still less likely to be caused by selection, increasing the strength of the evidence for neighbor-influenced mutation rates.

Returning to the dynamic aspect, we note from equation (9) that marked linkage disequilibrium will arise between the segregating sites if $N_e c$, the product of the effective population number and the recombination fraction, is less than unity. For example, if $N_e c = 0.5$, we have $\sigma_d^2 = 3/13$ or $\sigma_d \approx 0.48$. Since σ_d is roughly equal to the correlation coefficient, this means that roughly 50% correlation exists between the frequencies of mutants at both sites.

According to NEI (1968), who listed the number of nucleotide pairs per unit map length for various organisms ranging from viruses to mammals, the recombination fraction between neighboring sites is about 4×10^{-8} for Drosophila and 4×10^{-9} for the mouse. These values are about the same order of magnitude as the mutation rate per nucleotide (cf. KIMURA 1968a,b).

Since the effective number (but not the actual number) of many species may not reach 10^6 and may be much less, it is expected that $N_e c$ is usually much smaller than unity for two segregating nucleotide sites within a cistron. This means that strong linkage disequilibrium is expected to be very common between segregating sites within a cistron. Although four "alleles" are theoretically possible by segregation in a single nucleotide site, it is much more likely that when three or more alleles are maintained in the population, two or more sites are involved. In such a case, polymorphism involving exactly three alleles necessarily means linkage disequilibrium within a cistron.

The fact that σ_d^2 approaches 5/11 or roughly 1/2 rather than 1 at the limit of $N_e c = 0$ may be understood by noting that under a steady flux of mutations, the

mutant frequencies at two segregating sites are usually different, and $\sigma_d^2 = 1$ cannot be attained even when there is no recombination.

Another interesting point related to our present result is that σ_d^2 becomes approximately $1/(4N_ec)$ if $4N_ec$ is much larger than unity.

$$\sigma_d^2 \approx 1/(4N_ec) \tag{10}$$

For two segregating sites that are 0.1 map unit or more apart, this should give a good approximation for most natural populations. In this case, the contingency coefficient of mutant frequencies between the two sites is roughly equal to $1/\sqrt{4N_ec}$.

We have already shown in our previous papers that this approximation holds for the case of steady decay (OHTA and KIMURA 1969a). It also holds for the stationary state attained under recurrent mutation or overdominance (OHTA and KIMURA 1969b, 1970).

Thus the approximation formula (10) seems to be applicable quite generally to two segregating sites as long as $4N_ec$ is large.

We would like to thank Dr. J. F. CROW for reading the manuscript and making valuable suggestions.

SUMMARY

Linkage disequilibrium or nonrandom association of mutant forms between two segregating nucleotide sites in a finite population was studied using diffusion models, assuming that the number of nucleotide sites making up the genome is so large while the mutation rate per site is so low that whenever a mutant appears, it represents a mutation at a homoallelic site, i.e., a site in which no mutant forms are currently segregating in a population.—It was shown that under steady flux of molecular mutations in a finite population, if we measure the amount of linkage disequilibrium between two segregating sites by

$$\sigma_d^2 = E\{D^2\}/E\{x(1-x)y(1-y)\},$$

where D is the ordinary coefficient of linkage disequilibrium, and x and y are the frequencies of mutants at the two sites, then we have

$$\sigma_d^2 \approx (5+2R)/(11+26R+8R^2),$$

where $R=N_ec$ in which N_e is the effective size of the population and c is the recombination fraction between the two sites.—It was pointed out that if multiple alleles in a random mating population are generated through segregation at two or more nucleotide sites within a cistron, a strong linkage disequilibrium is usually expected between those sites, even in the absence of selection.

LITERATURE CITED

CROW, J. F., 1969 Molecular genetics and population genetics. Proc. 12th Intern. Congr. Genet. **3:** 105–113.

JOSSE, J., A. D. KAISER and A. KORNBERG, 1961 Enzymatic synthesis of deoxyribonucleic acid. VIII: Frequencies of nearest-neighbor base sequences in deoxyribonucleic acid. J. Biol. Chem. **236:** 864–875.

KENDALL, M. G., 1948 *The Advanced Theory of Statistics*, Vol. 1. Charles Griffin & Co. Ltd., London.

KIMURA, M., 1968a Evolutionary rate at the molecular level. Nature **217**: 624–626. ——, 1968b Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. **11**: 247–269. ——, 1969a The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics **61**: 893–903. ——, 1969b The rate of molecular evolution considered from the standpoint of population genetics. Proc. Natl. Acad. Sci. U.S. **63**: 1181–1188.

KING, J. L. and T. H. JUKES, 1969 Non-Darwinian evolution: Random fixation of selectively neutral mutations. Science **164**: 788–798.

NEI, M., 1968 Evolutionary change of linkage intensity. Nature **218**: 1160–1161.

OHTA, T., 1968 Effect of initial linkage disequilibrium and epistasis on fixation probability in a small population, with two segregating loci. Theoret. Appl. Genet. **38**: 243–248.

OHTA, T. and M. KIMURA, 1969a Linkage disequilibrium due to random genetic drift. Genet. Res. **13**: 47–55. ——, 1969b Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. Genetics **63**: 229–238. ——, 1970 Development of associative overdominance through linkage disequilibrium in finite populations. Genet. Res. **16**: 165–177.

APPENDIX

Basic equations for deriving the moments of the steady flux distribution. I. The moment equations for nonequilibrium population:

We will first consider the single-variable case. Let x_t be the frequency of a mutant form at a given site at time t (conveniently measured with one generation as the unit length of time), and let δx_t be the amount of change in x_t during a short time interval between t and $t+\delta t$ so that

$$x_{t+\delta t} = x_t + \delta x_t. \tag{A1}$$

Let $f(x)$ be a polynomial of x and consider the expected value of this function with respect to the frequency distribution at time $t+\delta t$, i.e.,

$$E\{f(x_{t+\delta t})\}.$$

Substituting (A1) in this expression, we have

$$E\{f(x_{t+\delta t})\} = E_\phi E_\delta\{f(x_t+\delta x_t)\}, \tag{A2}$$

where E_δ is the operator of taking the expectation with respect to the change δx, and E_ϕ is that of taking the expectation with respect to the frequency distribution at time t.

Expanding the right-hand side of (A2) in terms of δx_t and neglecting the higher-order terms containing $(\delta x_t)^3$ etc., we get

$$E\{f(x_{t+\delta t})\} = E_\phi\{f(x_t)+E_\delta(\delta x_t)f'(x_t)+\frac{E_\delta(\delta x_t)^2}{2!}f''(x_t)\},$$

or

$$\frac{E\{f(x_{t+\delta t})\}-E\{f(x_t)\}}{\delta t}=E\left\{\frac{E_\delta(\delta x_t)}{\delta t}f'(x_t)+\tfrac{1}{2}\frac{E_\delta(\delta x_t)^2}{\delta t}f''(x_t)\right\},$$

where E now denotes E_ϕ. At the limit of $\delta t \to 0$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}E\{f(x)\}=E\left\{\frac{V_{\delta x}}{2}f''(x)+M_{\delta x}f'(x)\right\}, \tag{A3}$$

where $M_{\delta x}$ and $V_{\delta x}$ are the mean and the variance, respectively, of the rate of change in mutant frequency per generation and are approximations to $\lim_{\delta t\to 0}\{E_\delta(\delta x_t)/\delta t\}$ and $\lim_{\delta t\to 0}\{E_\delta(\delta x_t)^2/\delta t\}$. For the stationary distribution, the left-hand side of equation (A3) vanishes and it reduces to the equation (A2′) of OHTA and KIMURA (1969b).

Extension of equation (A3) to the multivariable case is immediate. For n random variables $x_1, x_2, \ldots, x_n$, let $f \equiv f(x_1, x_2, \ldots, x_n)$. Then we have

$$\frac{\mathrm{d}}{\mathrm{d}t}E(f)=E\{L(f)\}, \tag{A4}$$

where L is the differential operator

$$L=\tfrac{1}{2}\sum_{i=1}^{n}V_{\delta x_i}\frac{\partial^2}{\partial x_i^2}+\sum_{i>j}W_{\delta x_i\,\delta x_j}\frac{\partial^2}{\partial x_i\partial x_j}+\tfrac{1}{2}\sum_{i=1}^{n}M_{\delta x_i}\frac{\partial}{\partial x_i} \tag{A5}$$

in which M, V, and W designate, respectively, the mean, the variance, and the covariance in the rate of change in the random variables that appear as subscripts.

II. *Steady flux case:* In the single-variable case, we assume that mutational input occurs at $x=p$ and output due to extinction or fixation occurs either at $x=0$ or $x=1$. At the state of steady flux, input and output balance each other and a steady state is reached with respect to the probability distribution, $\Phi(x)$, of mutants among segregating sites. Thus, writing $\Delta_{mut}E(f)$ for the input by mutation with respect to $E(f)$, we have

$$\Delta_{mut}E(f)+\frac{\mathrm{d}E(f)}{\mathrm{d}t}=0 \tag{A6}$$

or

$$E\left\{\frac{V_{\delta x}}{2}f''(x)+M_{\delta x}f'(x)\right\}+\Delta_{mut}E(f)=0. \tag{A6′}$$

It is important to note here that equation (A6) or (A6′) is valid only for $f(x)$ which vanishes both at $x=0$ and $x=1$, because the steady state distribution $\Phi(x)$ refers only to unfixed classes ($1/2N \leqq x \leqq 1-1/2N$), and new fixations that occur each generation at the terminal classes $x=0$ and $x=1$ should not be included in $E(f)$.

As an example of application of equation (A6′), consider the case of selectively

neutral mutations. We assume that each generation mutation occurs in the entire population at ν_m sites and that the effective population size is N_e. If we take $f(x)=2x(1-x)$, $H \equiv E(f)$ represents the number of heterozygous nucleotide sites per individual as considered by KIMURA (1969a). Since for this case $M_{\delta x}=0$, $V_{\delta x}=x(1-x)/(2N_e)$, and $\Delta_{mut}E(f)=2\nu_m p(1-p)$; noting $f'(x)=2(1-2x)$ and $f''(x)=-4$, we obtain, from equation (A6′),

$$-\frac{1}{2N_e}E\{2x(1-x)\}+2\nu_m p(1-p)=0$$

or

$$H=E\{2x(1-x)\}=4N_e\nu_m p(1-p). \tag{A7}$$

This agrees with the result obtained by KIMURA (1969a) using a different method.

The above treatment may readily be extended to the multivariate case, and we obtain

$$E\{L(f)\}+\Delta_{mut}E(f)=0, \tag{A8}$$

where L is the differential operator given by (A5).